© 2017 IJSRST | Volume 3 | Issue 8 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
Themed Section: Science and Technology

Analysis of Different Recognition Technique Classification and their Accuracy Measure over Odia Script-A Review

Rajat Ketan Sahu*, Anita Minz, Sagarika Mishra, Chandrasekhar Panda
Department of Computer Science and Application, Sambalpur University, Jyoti Vihar, Burla, Odisha, India
rajat.keten.sahu@gmail.com, dr.chandrasekharpanda@gmail.com

ABSTRACT

This paper presents a literature review on character recognition techniques. (OCR) Optical Character Recognition frameworks are able to identify varieties of text styles and printed characters. This means that by scanning a document the device is able to essentially 'read' the content. Character identification is a requesting specialization of pattern recognition, machine learning, and advance image processing. Various work has been done as such far for independent Indian language like Hindi, Urdu, Gujarati, Punjabi and onward. The latest research in this area has been capable to grow some new methodologies to overcome the difficulty of Odia writing style. Be that as it may, narrow exploration has been done over on odia content. Thus far various techniques are projected for segmentation, feature extraction, classification, and recognition tasks of Odia characters. This paper tries to review, their relative strength and weakness unit layout done by various research worker over on Odia content in last few years. The basic objective of this paperwork is to nearby present an overview of different existing methods that have been developed to recognize Odia character throughout last decade. The paper may be a concern the individuals who are interested to work in the fields of acknowledgment of Odia script.

Keywords : Optical Character Recognition (OCR), Odia Script, Classification, Recognition.

I. INTRODUCTION

Optical character recognition (OCR) deal with the technique of converting or changing over the handwritten text or printed content into a machine representation, that is understood by the machines for easy editing and looking. The greatest trail inside the field of image processing is to acknowledged the documents both in machine printed and handwritten structure [1]. Optical Character Recognition is a type of document image analysis where a scanned digital image that involves either machine printed or handwritten script input into an OCR software engine converts it into an editable machine readable digital text format[2].Development of OCR for different Indian languages is an active community of research work in today's era. We are creating and aim to develop the Hand written Character recognition framework for Odia conversation, which is the official language of Odisha.A part of research, is done on the printed Odia script, however a little mass of work has been done on the handwritten Odia content for recognition. Partition of handwritten or printed text into lines, words, and characters is one of the prime footsteps in the handwritten text recognition outlook. The text of hand written can be difficult to recognize due to variation in font sizes, type of noises, color, contrast, alignment, and background information. This kind of modification is difficult to identify the word in the document while writing [3]. Since handwritten text will vary on the user's skill, orientation and manner of the writing method. An in-depth study of Odia character recognition, outlining existing ways and their drawbacks have been carried out in this paper and future research direction in this field has been suggested. The various steps of character recognition such as image acquisition, preprocessing, segmentation, feature extraction, and post-processing.

There are collections of application domain where, OCR can help. Major areas are represented below:
1. Protect past records in electronic format.
2. Save document pictures inside limited space.

3. Facilitate visually impaired persons to scan the content of the document.

II. OCR METHOD

2.1 Image Acquisition:

Image acquisition may be a method the picture is through a camera or some scanner and feeding it to the computer for additionally processing. The images are represented within the format similar to JPEG, BMT, TIF, and TNG. The input image is also gray, color [4] [14].

2.2 Preprocessing:

Pre-processing deals with improving quality of image for better recognition by the system. The fundamental goal of pre-processing is to supply information that may be simply worked for the OCR system. The objectives of pre-processing are Binarization, Noise reduction, Scale, and normalization Aspect ratio, Skew correction, line removal etc. [5].

2.3 Segmentation:

The Segmentation separates a text document into line, word, and character [6]. Content Line division depends on (Hough Transform, Horizontal projections, smearing), Word partition is based on (vertical projections, connected element analysis), Character segmentation is predicted on (Vertical projections, Feature Extraction) etc.

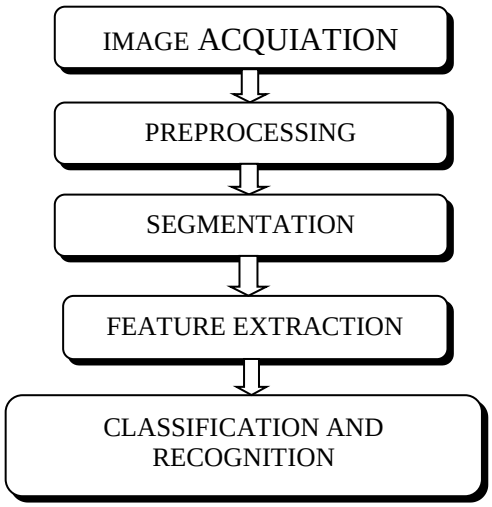


Figure-1: (STAGES OF OCR)

2.4 Features Extraction:

The process of extracting the features from object/alphabet to form a feature vector, which becomes its identity. Features perform necessary part in handwriting recognition system. The major purpose of feature extraction is to obtain a set of features, which maximizes the recognition rate with effectively representation of elements [7].

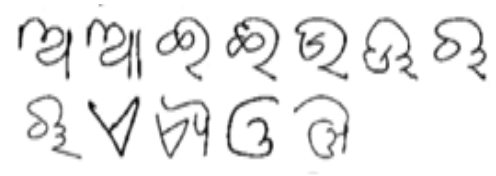
2.5 Classification:

Classification is the final stage of character recognition task to optimize the whole recognition process. The classifier measures the input with the stored feature determine a class. Universally used classifier are Template matching, Artificial Neural network(ANN), k-Nearest Neighbor, Byes Classifier, Neural Networks, Hidden Markov Models (HMM), Support Vector Machines, etc.The Neural network used to overlook the performance of feature sets[8].

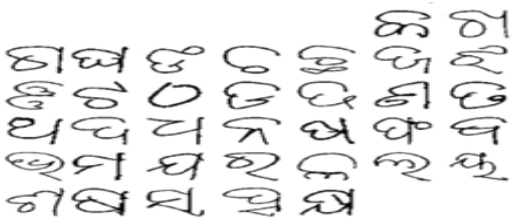
III. FEATURES OF ODIA SCRIPT

Odia script is utilized to compose the Odia language. The language is also spoken by minority populations of neighboring states, where local speakers mixed up 80% of the populace and rest 20%originates from different parts of West Bengal, Jharkhand, Chhattisgarh, and Andhra Pradesh. It is the official language of Odisha state and is the second language of Jharkhand. Odia is the sixth Indian language to be recorded as a classical language in India, on the premise of having a long abstract history and not having acquired broad topic from different dialects. Odia is a language used in the state of Odisha, India. The odia script created from the Kalinga script that is descendent of Brahmi script of ancient India [4,9,14]. The modern Odia content has 12 vowels and 35 consonants and ten Odia numerals. These are called the basic characters in Odia content[4]. Composing of the Odia script is from left to right and the idea of uppercase and easier case letters is truant. The characters are roundish in nature and look similar to each other. That makes it produces difficult for clustering. However, there are no flat lines like in Bengali and Hindi scripts which makes the division a ton less demanding[10].At the point when a vowel takes after a consonant, it takes an adjusted shape. Depending

upon the vowel, its changed shape is put at the left, right, best or at the base of the consonant. The adjusted shapes are called Modifiers or Matras. Also when a consonant or vowel takes after the consonant, it off and on again takes a changed shape known as a compound character [9][11].



Odia vowels



Odia consonants

Figure-2: Basic handwritten characters of odia alphabet [23].

A text in Odia scripts mainly divided into three zones: upper zone, the middle zone, and lower zone. The segment lying between mean line and higher line represent the upper zone. The middle zone consists of the area below the mean line and above the base line. The segment of the base line and lower line contains a lower zone where some of the modifiers are placed. The imaginary line isolating the center and lower zone is known as the base line. The line which separates the upper zone from the middle zone is named the mean line. Examples of zoning are appeared in Fig.3. in this case, the mean line along with a base line divided the text line into three zones[12].

Figure -3: Identification of different zones and lines of an Odia script [4].

IV. REVIEW OF LITERATURE

Chaudhari.B.et.al, 2002 in the review of Automatic recognition of printed Odia script is mentioned. The paper discusses the event of OCR for the script is difficult because a large number of character shapes in the script have to acknowledge. The no. of character is massive and two or more characters may combine or more to complex shape character referred to as compound characters. In this method, individual Characters are recognized using a combination of stroke &run based feature, along with feature obtained the concept of water flows from the reservoir to detect a vertical stroke of middle zone height. Classification depends on the choice of tree classifier. The advantage of this technique is simple & strong, and do not search for preprocessing steps like thinning and pruning. It is concluded that character segment need improve large touching character, which occurs in an inferior quality material. The work provides a mostly recognition accuracy of 96.3% [13].

Panda.S.R.et.al, 2015 Come up with the strategy on the advancement of the algorithm for Odia typewritten character recognition utilizing Template with Unicode mapping. In these work a database is created of pixel size 50*50.In the pre-processing grayscale image is converted to a binary image by selecting threshold value is called binarization for computing the matching metrics result. Classification is done with a set of prototype presenting each possible classes for Unicode mapping. Unicode standard features that each character code has a width of 16 bits. It is closed with that the calculation effectively tried and accomplished the exactness of 97.87% if there should arise an occurrence of Odia characters. Its work is limited to typewritten and enhanced towards handwritten [14] .

Rushiraj.I.et.al, 2016 reflected on their study that on handwritten character recognition of Odia script. attempted to classify 36 basic Odia consonants using 48 geometric features, were extracted by shadow features, centroid features and distance based features extracting geometric feature from individual character and comparing them with library feature by finding weight Euclidian distance for classification advantage of this method the Euclidean distance between two feature vectors measures the performance of similarity as it involves statistical analysis of training set features. Since the discriminative element set, is considered into account as the most critical factor to accomplishing the high recognition performance. Unicode is chosen for

encoding scheme for 16 bits character length and thus a way to recognition and transmission of characters simply. It is come to an end that OCR is highly dependent on the quality of input; it is not possible to compare to match among fully totally different techniques in terms of performance. The validity of the technique is 87.6% [15].

Nayak.M.et.al, 2015 found that odia conjunct character recognition using an Evolutionary algorithm. Most efforts have been devoted to the recognition of isolated and handwritten basic fonts. This paper introduces an effective text recognition scheme for a odia conjunct characters. Preprocessing require removing noise, skew, variation in structure of image, it converts the gray scale into binary image. Binary representation of symbol as feature ,binary symbol is resized (scaled)to a specific size of 15*15.here we compare two classifiers for conjunct character ,BPNN,ANN with weight using GA.advantage of this method it can find the near global solution in a large solution space quickly as of its parallel adaptive nature based on natural genetic system. Each gene of a chromosome represents a neural network. It is concluding that GA can be experimented with various cross over for conjunct characters of odia language, yields good results. Next adaptive versatile changing of the crossover probability may helps further improve efficiency segmentation may more improved which reduce the no. of training sets. The overall accuracy of method is 95.9% [16].

Mohanty.S.et.al, 2011 described an approach using bilingual documents that handles both Oriya and Roman script. Through that it calculates perfection for Odia alphabet with completely different font and size. It's far concluded with selection of features and designing of classifiers impact on accuracy by taking into different kinds of picture used in characters for improve over all the classification performance. Bilingual OCR handling degraded, noisy machine printed and italic textual content. Work may be extended similarly handwritten text [17].

Mishra.S.et.al, 2010 planned associated Odia character recognition utilizing neural networks. During this paper preprocessing technique is employed the image into smallest part elements with layout to find out text, line, words, character blocks. Character blocks compared to large dictionary of characters from different font style and language. Neural network totally based on feature matching. the approach used here is AFFNN, extracted geometric such as height, width, number of pixel and textual feature such as histogram.Advantagecentroid get a valid epoch or score, if the scored within desired range then character is recognized, else suggests for back propagate with overall score. It is concluding that due to no availability of 100% noise freed image and presence of similar shaped characters the accuracy rate is affected [18].

Basa.D.et.al, 2012 in this review explained the concept of odia handwritten character recognition using ANN. This paper discusses the character modeling; in pre-processing operation method we have to acknowledge a text in scanned document. In segmentations handwritten text is split into Line segmentation region unit at that point separated into word segmentation and word segmentation into character division. Character Recognition is based on method that uses the unique structure of some characters has found better result as compared to alternative strategies is mentioned throughout the paper. It is concluded that the recognition rate is very much affected by a similarity of different characters. Similar characters will degrade the identification rate [19].

Mohanty.S.et.al, 2011 in this paperwork, discusses the characteristics of some classification approach such as Support Vector Machines and K-nearest Neighborhood have been deal with Oriya characters. An object is classified by the "distance" from its neighbors with the object being appointed to the class most typical among its k distance. it is usual to use Euclidean distance to feature space. Kernel methods, including SVM are combining overall training and execution. SVM using multiclass problem that use linear kernel become simply and quickly. Advantage SVM has been superior classification accuracies to K-NN classifier in much experiment. When training through samples SVM give higher accuracies than statistical classifiers. It concludes that biggest limitation of the Support Vector approach lies in the choice of kernel. Second is in speed and size, both in training and testing. the overall accuracy is mentioned in SVM & KNN are 98.9% and 96.47%[20].

D.Padhi.et.al, 2012 This article focused a unique hybrid approach for Odia Handwritten character. Gives a general talk of standard deviation and zone centroid based feature preparing and testing utilized as a part of neural network that helps for recognition of offline character. This work is performed on Genetic Algorithm that helps optimum feature extraction and recognition. With the goal that technique for Odia HCR plays an important role to use a unique and strong combination of ANN and GA provides higher efficiency on programming and testing function module [21].

Tripathy.Jyotsnarani.et.al, 2010 proposed a method for Reconstruction of Oriya Alphabets Using Zernike. During this paper preprocessing technique used the document image is binarized, next extract the feature by Zernike moments technique, measured of pixel element around the center of gravity, then reconstruct image. moment based mostly in variation explore data across the entire image.Hu's moment variation of central moment compared with Zernike moment to achieve scale normalization, then the translation normalization moving the image center to map a unit circle. The complex polynomial, complete the orthogonal set over disk. it is concluding that the real and imaginary part together from the magnitude of the Zernike moments obtained the reconstruction of image[22].

Table I. Survey of Various Character Recognition and Their Accuracy for Odia Script

Title of the Paper	Author and Publication Details	Classification Technique	Recognition Technique	Recognition Accuracy rate	Limitation
Automatic Recognition Of printed Oriya script	Chaudhary.B.et.al [13]	Tree classifier,verticalStrokes, holes, Water flow Reservoir	Topological Feature, Combination of stroke and run Number based feature, Water flow reservoir	96.3%	Character phase improve large touching character in Inferior material, Boundary tracking technique for distinguish between similar character.
Odia Offline Typewritten Character Recognition Utilize Template Matching with Unicode Mapping	Smruti Rekha Panda et.al[14],	Template Matching, distance among pattern and class of prototype	Template coordinating with Unicode mapping ,vector matrix	97.87 %	Handwritten character and compound character acknowledge is difficult due to its variation in font and size Involvement of various noise and persons. Highly depend on quality of input makes difficult to evaluate and compare completely different system.

Handwritten character recognition of Odia script	Indugurushiraj[15],	Euclidian distance technique, feature library, Unicode mapping	48 Geometric feature, shadow, Centroid, distance base.	87.6 %	OCR is highly dependent on the standard of input hence it is not possible to check among totally different techniques in terms of performance.
Odia conjuct character recognition Using Evolutionaryalgorthim	Nayak.M.et.al [16]	Back -propagation neural network(BPN),Genetic neural network(GAONN)	Binary representation Symbol as vector.	95.9%	Adaptive ever changing the crossover probability mighthelp further improve efficiency and segmentation can also further improved which reduce the number of training sets.
A Comparative survey of Classifiers accuracies for Bilingual Printed Documents(Oriya-English)	Sanghamitra Mohanty.et.al [17]	K-Nearest Neighbor(KNN), Convolution Neural Network(CNN) And Support Vector Machine(SVM)	Fractal based texture features, Gabor filter		Can be extended To handwritten Content and grater refinement of bilingual OCR machine written text with least noise.
Oriya Character Recognition Using Neural Networks	Mishra.S.et.al [18]	Feed Forward Neural Network(FFNN), Back propagation	Dimension measurement,height, width, Centroid of Character coordinates		No availability of 100% noise freed image and presence of similar formed Characters, the accuracy rate affected.
Offline Character Recognition System using Artificial Neural Network	Nisha Vasudeva et.al[19]	Bayesian network classifier, Probabilistic methods	Model using mean square error and mean absolute error throughout ANN training	96 %	Efforts to try and made in obtaining higher accuracy to enhance recognition accuracy.

Performance Comparison Of SVM and K-NN For Oriya Character Recognition	Mohanty.S.et.al[20]	Support Vector Machine(SVM), K-Nearest neighborhood (K-NN)	Feature Weigh on basis longest run option features, Wrapper primarily Based feature.	SVM-98% KNN- 96.7%	SVM lies within the selection of Kernel; the Second limitation is in speed and size, both in training and testing.
Novel Hybrid approach for Odia Handwritten Character Recognition System	DebanandaPadhi et.al [21]	Hybrid approach of BPNN and GA	Fractal based mostly texture features options, Gabor filter.	94%	Work to be done on include feature extraction of the Compound character of manually written Odia style.
Reconstruction Of Oriya Alphabets Using Zernike Moments	TripathyJyotsnarani [22]	Zernike moments of the Orthogonal polynomial.	Hu seven moments, translation, Scale, rotation invariant.		Computation of Zernike moment is more complicated for image normalization Process.

Image Type	Size of samples	Accuracy percentage
ଅ ଆ ଇ ଈ	***Bold and small***	***92.78%***
ଅ ଆ ଇ ଈ	***Bold and big***	***98.9%***
ଈ ଉ ଉ ଋ	***Normal and small***	***96.98%***
ଅ ଆ ଇ	***Normal and bold***	***97.12***

Figure-4: Effect on accuracy by considering the different type of the images used for odia characters [17].

V. CONCLUSION

In this paper, we have presented a related work on Odia character recognition techniques. The analysis work assigned throughout the last decade among within in the field of Odia character. Various available techniques are studied to find out the foremost effective technique. Each of this strategy has its own benefits and limitations. But is found that the techniques which provide better results are slow in nature while fast techniques mostly provide inefficient results. Recognition of character continues to be a difficult task since there is a variation inside a similar character because of its various text dimension such as font size, different types of noises and involvement of various persons. However, some work has been reported on Odia character recognition. The area of character recognition in Odia language still wants an in depth-study. Numerous kinds of degradation found in Odia content, particularly in old books. Degradation may due to imperfection in papers. Two kinds of degradation touching characters and heavily printed characters. Furthermore, analysis work is needed to develop the performance of our system.

VI. REFERENCES

[1]. Vijendra, S, Vasudeva, N, &Parashar, H. J. (2011). Recognition of Text Image Using

Multilayer Perceptron. RESEARCH GATE, (Icmlc), 547-550.

[2]. Tawde, G. Y., &Kundargi, J. M. (2013). An Overview of Feature Extraction Techniques in OCR for Indian Scripts Focused on Offline Handwriting. International Journal of Engineering Research and Applications, 3(1), 919-926.

[3]. Tigga, D., Mishra, S., & Panda, C. S. (2017). Review on Offline Odia Character Recognition. International Journal of Computer Systems, 4(2), 24-29.

[4]. Pushpalata, P.,& Babita , M. (2015). A Survey on Odia Character Recognition. International Journal of Emerging Science and Engineering, 8(2), 2319-6378.

[5]. Sahu, A., & Mishra, S. (2015). Study and Analysis for Development of an Efficient OCR for Printed and Handwritten ODIA Documents : A Survey. Ijarcsse, 4(11), 14-16.

[6]. Jindal, P. (2015). Line and word segmentation of handwritten text documents written in Gurumukhi script using midpoint detection technique. Ijarse, 4(1), 11-19.

[7]. Sahu, M. S. and N. (2016). A survey on handwritten character recognition (hcr) techniques for English. Advance in vision computing: an international journal (avc), 3(1).

[8]. Goswami, M. M., Prajapati, H. B., &Dabhi, V. K. (2011). Classification of printed Gujarati characters using SOM based k-nearest neighbor classifier. ICIIP 2011 - Proceedings: 2011 International Conference on Image Information Processing.

[9]. Science, C. (n.d.). Optical character recognition Bachelor of Technology optical character recognition Bachelor of Technology, 1-28.

[10]. Chaudhuri, B. ., & Pal, U. (1998). A complete printed Bangla OCR system. Pattern Recognition, 31(5), 531-549.

[11]. Pal, U., Wakabayashi, T., & Kimura, F. (2007). A system for off-line oriya handwritten character recognition using curvature feature. Proceedings - 10th International Conference on Information Technology, ICIT 2007, 227-229.

[12]. Tripathy, N., & Pal, U. (2004). Handwriting segmentation of unconstrained oriya text. Proceedings - International Workshop on Frontiers in Handwriting Recognition, IWFHR, 31(December), 306-311.

[13]. Chaudhury, b., pal, u.,&mitra,m. (2002). automatic recognition of printed oriya script.research gate, 27(january), 23-34.

[14]. Panda, S. R. (2015). Odia Offline Typewritten Character Recognition using Template Matching with Unicode Mapping. ISACC, 1-7.

[15]. Rushiraj, I., Kundu, S., & Ray, B. (2016). Handwritten Character Recognition of Odia Script. International Conference on Signal Processing, Communication, Power an.

[16]. Nayak, M.,& Nayak, A. K. (2015). odia-conjunct character recognition using evolutionary algorithm. asian journal of applied science, 3(4), 789-798

[17]. Mohanty, S., Nandini, H., &Bebartta, D. (2011). A Comparative Analysis of Classifiers Accuracies for Bilingual Printed Documents. International Journal of Computer Science and Information Technologies, 2(2), 916-923

[18]. Mishra, S., Nanda, D., &Mohanty, S. (2010). Oriya Character Recognition using Neural Networks. IJCCT, 2(2,3,4), 88-92

[19]. Vasudeva, N., Parashar, H. J., &Vijendra, S. (2012). Offline Character Recognition System Using Artificial Neural Network. International Journal of Machine Learning and Computing, 2(4), 4-7.

[20]. Mohanty, S., Nandini, H., &Bebartta, D. (2011). Performance Comparison of SVM and K -NN for Oriya Character Recognition. (IJACSA) International Journal of Advanced Computer Science and Applications, 112-116.

[21]. Padhi.D,(2012) "Novel Hybrid approach for Odia Handwritten Character Recognition System", International Journal of advanced Research in Computer Science and SoftwareEngineering,2(5).

[22]. Tripathy, J. (2010). Reconstruction of Oriya alphabets using Zernike moments. International Journal of Computer Applications, (0975-8887), 8(8).

[23]. Mohapatra, R. K., Mishra, T. K., Panda, s., &Majhi, B. (2016). ohcs: a database for handwritten atomic odia character recognition. in 2015 5th national conference on computer vision, pattern recognition, image processing and graphics, ncvpripg 2015.